MSDS 430: Python for Data Science

Final Project

Northwestern University, Masters of Science in Data Science

Predictive analytics plays an enormous part in today's digital marketing field. With the rise of internet technologies and big data, online companies are given the tools to collect and analyze information about their customers and the customers' product preferences. Data manipulation and analysis help deliver valuable insights about the performance of digital marketing campaigns and its success in terms of deliverability, target audience, product preference, the right combination of time, place, and target market.

My final project aims to follow the above logic. I have chosen to analyze a set of data containing information about the performance of social media online ad campaigns. With the help of Python, my goal is to reveal the best performing campaign in terms of impressions, clicks, spent, and conversions. I planned to explore which demographic groups responded best to the campaigns, how much did they spend, and what was the conversion rate. The purpose of the results is to be better informed for further building of new ad campaigns that are better targeted and will have greater success.

Data source: https://www.kaggle.com/loveall/clicks-conversion-tracking

I started by using df.head function to quickly preview the information the data set contained. Further I used df['age'].value_counts() and df['gender'].value_counts() functions to evaluate the representation of gender and age groups. The findings will guide the optimization process of the campaigns in the future. I found that the most active shopping age group was 30-34 yrs old, while males and females were very close in participation. Next steps were to investigate which ads and campaigns had most spend, conversions using sum calculations - xyz_df = df.groupby("xyz_campaign_id")["Approved_Conversion"].sum().sort_values(ascending=False) xyz_df.head() and xyz_df = df.groupby("xyz_campaign_id") ["Spent"].sum().sort_values(ascending=False)

xyz_df.head(). I used sum calculations to find out the most spend and conversions by interest group, age group, and gender group. To discover additional correlations, I used a pandas data frame. It guided me to further see correlations in the data with the help of pandas, matplotlib, and seaborn. I created a few scatter plots, box plots, bar charts that visually demonstrate the spending and conversions per person, the correlation between impressions and conversions, gender, age and spending, gender, age and interest, and finally, the ad and campaign with most conversions. The analyzed data gives informative results that could potentially help a marketer decide how to optimize the current campaigns for optimal results in the future, depending on the business goal of the company.
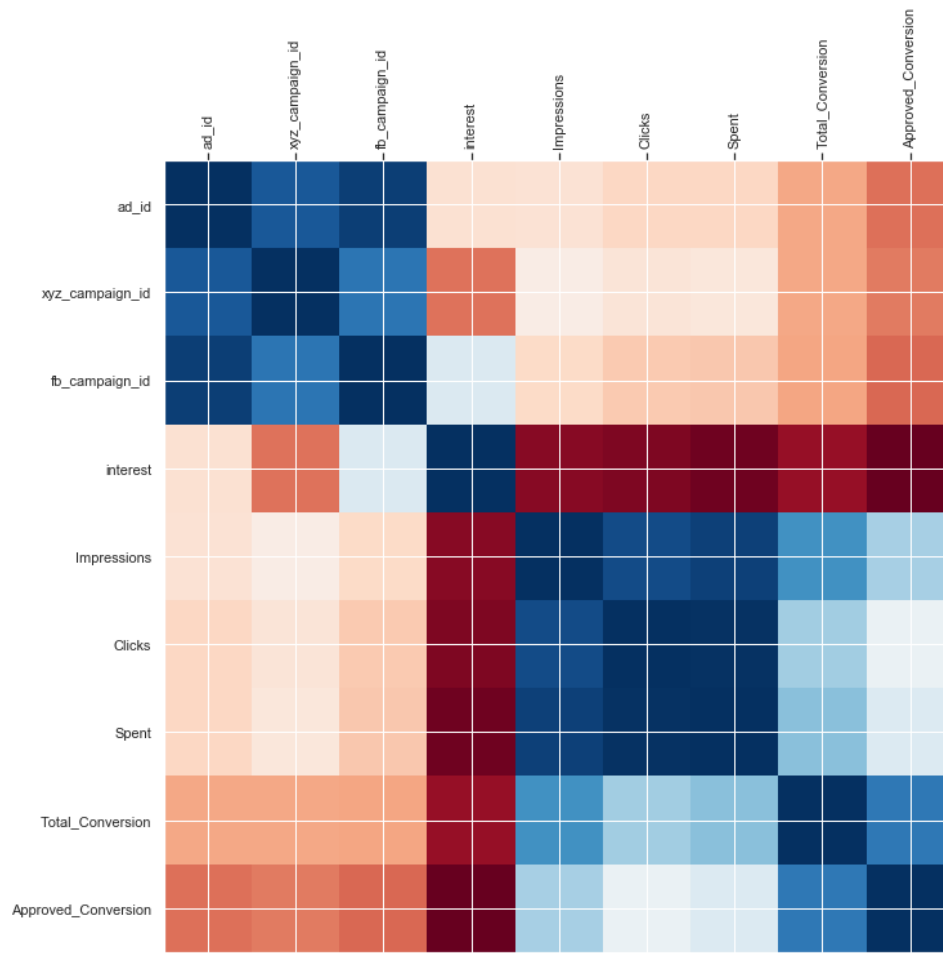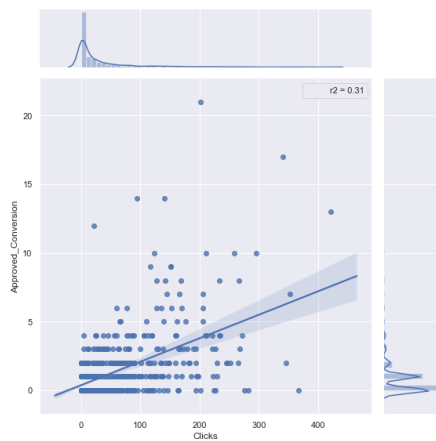
The average person had either 0 or 1 conversions, anything over 2 is probably an outlier.
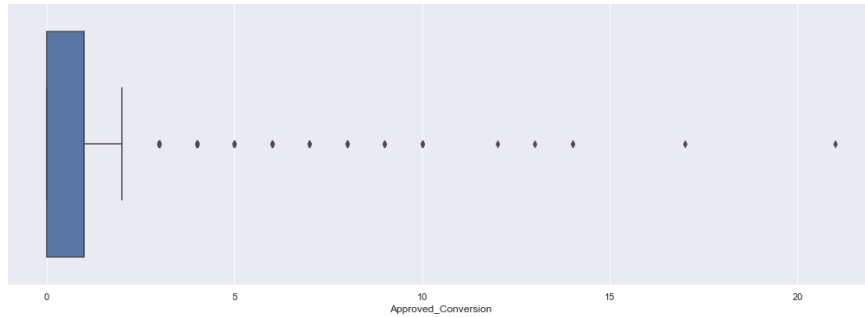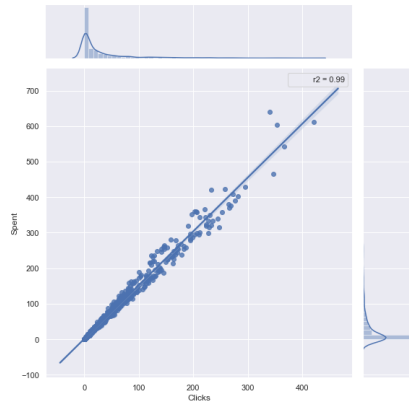

Average spending between 0 and 100 dollars, several outliers though
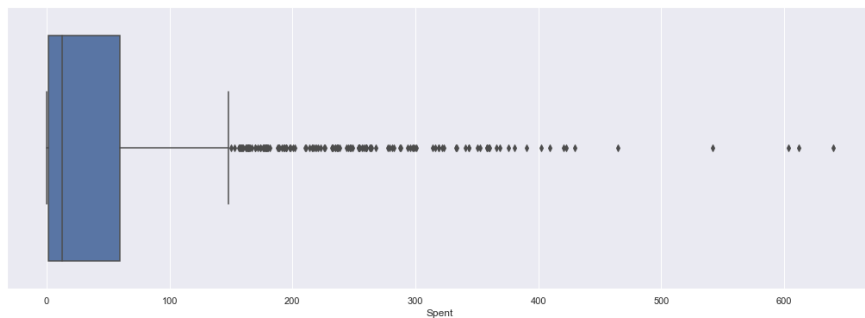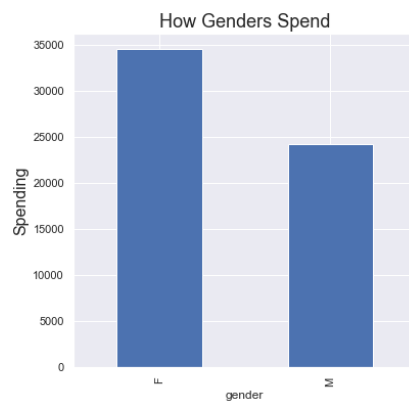

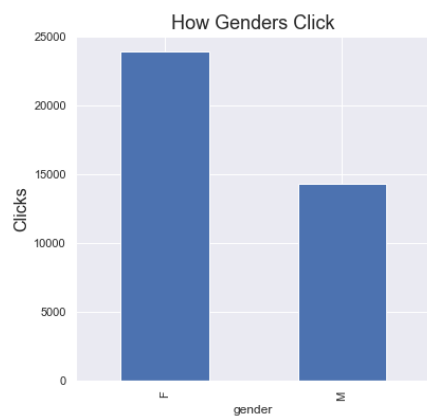**Clicks did not correlate well to conversions (lots of window shopping)**

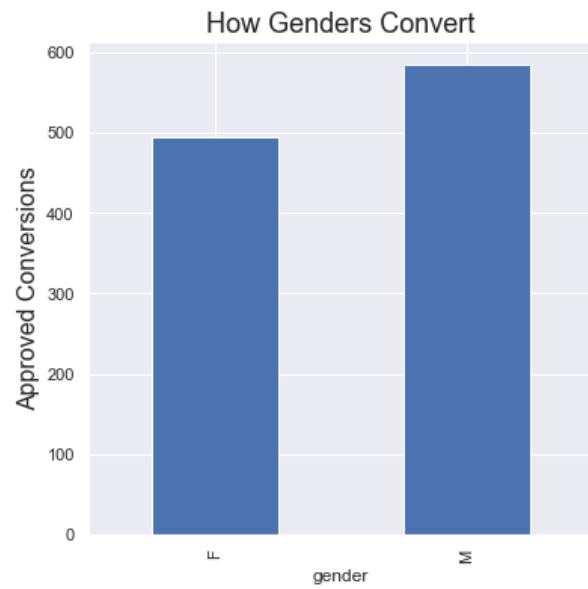**However, clicks did strongly correlate positively to spending**

Graph showing females spending more than males

Females click ads more than males

Males had more approved conversions.

How Genders Convert

Age group 45-49 spent most.

Spending per Age Group

Conversions per Age Group

Clicks per Age Group

Conversions per Campaign Id